9/29/2009

NYSERDA CFL Expansion Plan Evaluation – 2009 Baseline RDD Survey

The purpose of the NYSERDA CFL Expansion Plan Evaluation is to furnish information on the effectiveness of the CFL Expansion Plan. The purpose of the 2007 Baseline RDD Survey is to furnish benchmarks for New York State, New York City, and comparison areas – Washington, DC, Houston, TX, and Ohio – prior to the start of the CFL Expansion Plan.

Sample

Target Population

The target populations for the survey are telephone households in New York State (excepting New York City and Long Island), New York City, Washington, DC, Houston, TX, and Ohio.

Sample Frame and Selection

The RDD sample frames for each of the five locations were provided by Survey Sampling International (SSI). The New York State frame included all counties of New York excepting those on Long Island and in New York City. The New York City frame included the five boroughs of the City. The Washington, DC frame included only the District of Columbia. The Houston frame included all of Harris County. The Ohio frame included the entire state excluding area codes 513 and 283. These area codes cover the Duke Power service territory where a CFL program is being implemented.

The random digit sampling procedure provides representation of both listed and unlisted (including not-yet-listed) numbers by random generation of the last two digits of telephone numbers selected on the basis of their area code, telephone exchange (the first three digits of a seven digit number), and bank number (the four and fifth digits). Telephone exchanges are selected with probabilities proportionate to their size by county and by exchange within county. Only working banks of telephone numbers were selected. A working bank is a group of 100 contiguous telephone numbers that contain at least one working residential listing. Using working banks improves the efficiency of the sample in locating households and it provides the opportunity for households with numbers in a new exchange to be included in the sample.

The sample was released for interviewing in replicates, which are random subsamples of the larger sample. Using replicates controls the release of sample to ensure that all released sample numbers receive the full call procedures and to maintain appropriate regional distribution of called numbers.

Data Collection

Overview of Data Collection Procedures

The RDD Survey was administered as a telephone interview with a household. Interviewers from Braun Research conducted the interviews using a computer-assisted telephone interview (CATI) survey instrument.

Survey Instrument

The survey instrument gathered information on respondents' awareness of CFLs, their usage of CFLs in and around their homes, and their CFL buying habits. The instrument also addressed other household lighting devices, namely light emitting diodes (LEDs) and standard incandescent bulbs. The survey was pretested with a small number of respondents to evaluate the degree to which questions were clear and understandable and to check the logic and skip patters of the CATI survey instrument. A second Spanish-language instrument was also created and administered in New York City and Houston, TX.

Survey Administration

Interviewers had a minimum effort requirement of at least two daytime, evening, and weekend phone calls per sample telephone number. If the interviewer reached an answering machine at the number dialed, he/she left a message on the first contact and on every third contact thereafter. If the interviewer reached a household in New York City or Houston, TX that only spoke Spanish, the Spanish-language interview was conducted. Fielding began about two weeks prior to the holiday season, and was suspended for several days including Christmas Eve, Christmas Day, Dec. 26th, and New Year's Day. The New York State study was fielded for 37 days, New York City for 30 days, and the Washington, Houston, and Ohio studies were fielded for 25 days. Survey administration averaged 16 minutes per completed interview for all five studies.

Tables 1a to 1e show the final disposition of the samples.

Table 1a – Survey Sample Disposition: New York State

Disposition		**Number**	**Percent**
Complete	Complete	1001	25%
	Partial	56	1%
Contacted	Refused	1418	36%
	Not Completed	578	15%
Not Contacted	Quota Met	0	0%
Excluded	Duplicate	0	0%
	Contact no longer available	0	0%
	Information not available for contact	0	0%
	Unusable number	915	23%
	Not Eligible	2	~0%
Total		**3970**	**100%**

Table 1b – Survey Sample Disposition: New York City

Disposition		Number	Percent
Complete	Complete	502	21%
	Partial	33	2%
Contacted	Refused	867	36%
	Not Completed	509	21%
Not Contacted	Quota Met	0	0%
Excluded	Duplicate	0	0%
	Contact no longer available	0	0%
	Information not available for contact	0	0%
	Unusable number	489	20%
	Not Eligible	0	0%
Total		**2400**	**100%**

Table 1c – Survey Sample Disposition: Washington, DC

Disposition		Number	Percent
Complete	Complete	500	21%
	Partial	41	2%
Contacted	Refused	706	29%
	Not Completed	174	7%
Not Contacted	Quota Met	0	0%
Excluded	Duplicate	0	0%
	Contact no longer available	0	0%
	Information not available for contact	0	0%
	Unusable number	979	41%
	Not Eligible	0	0%
Total		**2400**	**100%**

Table 1d – Survey Sample Disposition: Houston, TX

Disposition		**Number**	**Percent**
Complete	Complete	503	21%
	Partial	43	2%
Contacted	Refused	700	29%
	Not Completed	212	9%
Not Contacted	Quota Met	0	0%
Excluded	Duplicate	0	0%
	Contact no longer available	0	0%
	Information not available for contact	0	0%
	Unusable number	939	39%
	Not Eligible	3	~0%
Total		**2400**	**100%**

Table 1e – Survey Sample Disposition: Ohio

Disposition		**Number**	**Percent**
Complete	Complete	501	21%
	Partial	31	1%
Contacted	Refused	733	31%
	Not Completed	306	13%
Not Contacted	Quota Met	0	0%
Excluded	Duplicate	0	0%
	Contact no longer available	0	0%
	Information not available for contact	0	0%
	Unusable number	829	34%
	Not Eligible	0	0%
Total		**2400**	**100%**

Table 2 shows the number of sampled cases, the number of completed interviews, and the response rate by sample stratum.

Table 2 – Nonresponsive Referred Survey Response Rate

	New York State	New York City	Washington, DC	Houston, TX	Ohio
Number of Interviews	1001	502	500	503	501
Eligible Sample Size[1]	3053	1911	1421	1458	1571
Response Rate	32.8%	26.3%	35.2%	34.5%	31.9%

Data Processing

Coding

The survey instrument included 22 field-coded questions, 21 open-ended questions, and 4 closed-ended "Other (Specify)" questions. For the field-coded questions, pre-coded response options were visible on the CATI screen and available to the interviewer to code but they were not read to the respondent. The interviewer had the choice of either coding the response into one or more of the pre-coded response options or coding the response as "Other" and recording the respondent's answer verbatim. For the open-ended questions, the interviewer recorded the respondent's verbatim response to the question. For the close-ended "Other (Specify)" questions, the interviewer asked a question with pre-coded response options but also had the ability to code the response as "Other" and enter a text string to summarize the response if it fell outside of the responses listed.

For each applicable question, APPRISE staff reviewed the verbatim response and then selected one of the pre-coded responses or created a new code if there were enough similar verbatim responses. A new code was created if 1% or more of respondents provided the same answer or if a new code was deemed useful for inclusion in future RDD evaluations. After reviewing the verbatim responses from the New York State study, the first site to be completed in the field, verbatim responses were grouped into categories and coded. A standard set of codes was developed for the New York State study then submitted to Nexus Market Research for review. Following approval, the same codes were used for the four following studies.

Data Processing

The survey data were checked for consistency with the CATI survey instrument, all skip pattern logic in the instrument was checked, and additional data consistency checks were run on the survey data. The survey data were combined with sample frame data, including respondent zip and county. Data were

[1] Eligible sample size is calculated by adding the number of eligible respondents to the number of cases where eligibility was unknown multiplied by the estimated eligibility rate.

delivered in a number of data file formats, including SAS, SPSS, Stata, and Excel. All files were labeled with variable labels and value labels.

The survey data was divided into three deliverable data sets: a CFL retailer data file, a CFL saturation data file, and an RDD data file. The three sets included the responses to different sets of questions. The CFL retailer data included questions BUY10 to BUY11-2 that asked the name and address of retailer where the respondent purchased CFLs. The CFL saturation data included question REC3 to that asked the respondent for their town, name, and phone number for participation in a further study. The RDD data contained the responses from all the remaining questions in the study.

Additional data processing was done with the RDD data for the New York State and Ohio studies. In these studies, respondents were asked to name their electric utility provider, and if they were unable to identify their provider, the town or city they live in. This information was checked to verify that the information the respondent provided was accurate, then both the original responses and corrected responses were included in the data file. This correction process began first by back-coding all the "other specify" responses, where appropriate, into existing response categories. Second, respondents who identified municipal utility providers or a rural electric association (REA, Ohio study only) were identified in the remaining verbatim responses. These utility providers were categorized in the corrected data as Municipals and REAs. Third, all responses were checked to confirm that the utility the respondent provided does in fact serve the location of their residence. In the New York State data, a listing of zip codes and electric utility providers that serve each zip was obtained from NYSERDA. Each of the New York State responses were then force cleaned against this list and re-categorized where there existed a discrepancy between the utility noted by the respondent and the utility serving the respondent's zip code. In the Ohio data, the utility noted by the respondent and the county in which they reside were checked against the service territory of each electric utility provider. For the twelve cases in which the utility provider given by the respondent does not serve the respondent's location, the response was re-categorized as a "Don't Know." A new variable was created in the data set (called "dem13_edited" that reflects these edits to the utility company question.

In all five studies, the BUY10 to BUY11-2 data was processed to highlight several facts: number of CFLs purchased by the respondents, the type of stores they were purchased from, the name of the stores, and location of the stores. This information was used to determine the number and percentage of households reporting purchasing CFLs by store name, type, and location. If a respondent said that they did not know how many light bulbs they bought, that number was replaced with the average of all the light bulbs bought by the respondents.

The CFL retailer data was cleaned by store name and location to determine the number of discrete stores where the CFLs were purchased. The data was first transformed to one record per store with each record containing basic information about the retailer. Then, an edited store name, address, city, and state was assigned to each of the retailers. This corrected information will be used for the CFL retailer survey.

Weighting

Since RDD surveys tend to under-represent certain parts of this target population (households containing only young adults, households with household heads with high school education or less, and single person households for example), relative weights were developed to furnish results that are

consistent with the geographic and demographic distribution of households in these areas. The population parameters used to calculate the weights were derived from the American Community Survey (ACS).

Weighting the data is an important means of mitigating the effects of differential non-response by geographic or demographic group and the inability of RDD samples to include households that are cell phone only – households that have no other telephone besides cell phones. Current data indicates that 14.8% of American households are cell-phone only households.[2] These cell phone only households tend to be more likely to contain younger adults and to be single adult households. The weighting is intended to more appropriately represent these types of households, balance the sample, and bring it into alignment with population parameters on geographic location, household member age, highest level of education by head of household, and household size.

The weighting for the New York State and New York City studies were calculated in two stages: a geographic weight and a demographic weight. These two weights were calculated separately but then combined into one final weight. The Ohio, Washington, and Houston studies included only a demographic weight. As cities, there was no way to subgroup the Washington and Houston studies to create geographic weights. For Ohio, an exploratory analysis of the rural versus urban areas of the state showed no significant differences in sample distribution from the actual population distribution and negated the need to create a geographic weight. The following procedures were employed:

- Geographic Weight: New York State and New York City Studies Only
    - Counties were combined into regions based on NYSERDA New York Energy Smart Community Regions.
    - The population parameter for the geographic weighting was developed using household counts by county from ACS 2005-2007 3 year average data. Schuyler and Hamilton counties were excepted from this process. Data for these counties is from 2000 Census.
    - Geographic weights were created to bring the sample distribution across regions into alignment with distribution of households across regions.
- Demographic Weight: All Studies
    - Demographic data on the ages of adult household members, highest level of education of the household head, and household size was obtained from ACS 2007 data.
- Complete Weight
    - The sample data was weighted with a geographic weight described above for only the New York State and New York City studies. The weighted sample distribution was compared on demographics (age, education, HH size) to population parameters. A demographic weight was calculated to bring sample distributions into alignment with

[2] CDC/National Center for Health Statistics. NHIS – Wireless Substitution, July December 2007. <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200805/htm>

population parameters. The weight at this stage is the geographic weight * demographic weight.

- The weights were trimmed so no individual case has too much impact on weighted data.

- An adjustment was calculated for use when New York City and State data are combined. This multiplier should be used only when analysis is based on the total New York State data (State and City data combined). The final New York City weight should be multiplied by 1.4848 and the final NY state weight should be multiplied by .7612 when these two data sets are combined and analyzed together as "total New York State (excluding Nassau and Suffolk counties)."